REMARKS

This application has been reviewed in light of the Office Action dated April 17, 2006. In view of the foregoing amendments and the following remarks, favorable reconsideration and withdrawal of the rejections set forth in the Office Action are respectfully requested.

Claims 13, 14, 16-24, 37, 38, 40-48, 50 and 52 are pending. Claims 13, 14, 16, 18-24, 37, 38, 40, 42-48, 50 and 52 have been amended. Support for the claim changes can be found in the original disclosure, and therefore no new matter has been added. Claims 13, 37, 50 and 52 are in independent form.

Claims 13, 14, 16-24, 37, 38, 40-48, 50 and 52 were rejected under 35 U.S.C. § 112, first paragraph, as containing subject matter not described in the specification in such a way as to enable one skilled in the art to make and use the claimed invention. Applicant respectfully traverses this rejection, for the reasons set forth below.

The claimed speech detecting means, on the one hand, and the claimed likelihood determining means and location determining means, on the other hand, are disclosed in the specification as (i) distinct from each other, and (ii) usable in combination with each other. The portion of the specification under the heading "Speech Detection" (pages 15-25) and the portion of the specification under the heading "Maximum Likelihood End-Point Detection" (pages 25-29) are not alternative embodiments but parts of a single embodiment. Specifically, both the "Speech Detection" portion and the "Maximum Likelihood End-Point Detection" portion of the specification are encompassed within the embodiment illustrated in Figs. 3 and 6a-9.

As described in the discussion in the specification pertaining to Fig. 7, the average signal energy per received frame is calculated, a sequence of energy values (representing a

sequence of frames) is filtered by bandpass filter 80, and modulation power calculation unit 82 calculates the modulation power of the filtered sequence. The bandpass modulation power for any given frame k, w_k, is compared with a detection threshold Th in threshold circuit 84, which outputs a control signal to control unit 86. (Assuming the apparatus is in the INSILENCE state,) the control signal indicates whether the bandpass modulation power w_k exceeds the detection threshold. When control unit 86 determines (as illustrated in Fig. 8a) that the bandpass modulation power w_k has exceeded the detection threshold for a predetermined number of frames (i.e., when control circuit 86 determines that CNTABV > NDTCT) (step S9 in Fig. 8a), control circuit 86 concludes that speech has begun (i.e., control circuit 86 concludes that speech has been detected), and control circuit 86 initiates a maximum likelihood calculation to determine (e.g., more accurately) the starting point of speech (step S28 in Fig. 8a). The maximum likelihood calculation is described in the specification at pages 25-29. As specifically stated in the specification, the maximum likelihood calculation is performed after speech detection has been made: "when the control unit 86 identifies that speech has started, it [begins the maximum likelihood calculation]" (page 25, lines 25*ff*; emphasis added).

With respect to Fig. 7, then, the claimed speech detection is performed using, e.g., elements 80, 82, 84 and 86, with element 86 ultimately indicating speech or non-speech. The claimed likelihood determination and location determination is carried out, e.g., by element 94. With respect to Fig. 8a (in which speech is being detected from the INSILENCE state), speech detection is first performed, with the actual detection of speech occurring at (or, based on the result of) step S9; if speech is detected at step S9, then the start of speech is detected, using a likelihood method, at step S28. See the specification at page 22, lines 14*ff*: "Once the count

CNTABV is above NDTCT, indicating speech has started, then the processing proceeds from step S9 to step S28, where the control unit 86 initiates the calculation of the start of speech point using a maximum likelihood calculation on recent frames" (emphasis added). Likewise, when the apparatus is in the INSPEECH state, non-speech is detected, followed by the calculation of the endpoint of speech using a likelihood method: "Once the number of consecutive frames below the threshold [CNTBLW] has exceeded NEND [i.e., once non-speech is detected; step S37, Fig. 8b (note: in Fig. 8b, "NHLD" should read "NEND" in step S37)], the processing proceeds to step S45 [Fig. 8b], where the control unit 86 initiates the calculation of the endpoint of speech using a maximum likelihood calculation with recent frames" (specification, page 24, lines 23*ff*; emphasis added).

Applicant submits that the claims comply with 35 U.S.C. § 112. In view of the above remarks, withdrawal of the rejection under Section 112 is respectfully requested.

Claims 13, 18, 21, 37, 42, 45, 50 and 52 were rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 5,638,487 (*Chigier*) in view of U.S. Patent No. 5,649,055 (*Gupta et al.*).

Claims 14, 22, 38 and 46 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Chigier* in view of *Gupta et al.* and further in view of U.S. Patent No. 5,842,161 (*Cohrs et al.*).

Claims 16, 17, 19, 20, 40, 41, 43 and 44 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Chigier* in view of *Gupta et al.* and further in view of U.S. Patent No. 4,956,865 (*Lennig et al.*)

Claims 23 and 47 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Chigier et al.* in view of *Gupta et al.* and *Cohrs et al.* and further in view of the article "Bounds on $R_1(D)$ Functions for Speech Probability Models" (*Abut et al.).*

Claims 24 and 48 were rejected under 35 U.S.C. § 103(a) as being unpatentable over *Chigier et al.* in view of *Gupta et al.* and *Cohrs et al.*, and further in view of U.S. Patent No. 5,778,342 (*Erell et al.*).

Without conceding the propriety of the rejections over the prior art, the independent claims have been amended. Applicant submits that the amended independent claims are allowable for at least the following reasons.

Independent Claim 13 recites, *inter alia*, speech detection means operable to process a received signal and to identify when speech is present in the received signal, means for determining a likelihood that a boundary is located at each of a plurality of possible locations within the energy signal, and means for determining the location of the boundary using the likelihoods determined for each of the possible locations, wherein the likelihood determining means is restricted to determine the likelihoods in the received signal only when the speech detecting means detects speech within the received signal. Each of independent Claims 37, 50 and 52 recites similar or identical features. Applicant submits that nothing in the cited art would teach or suggest at least these features of the independent claims.

The invention claimed in the independent claims thus involves two-stage speech detection, in which the second stage may be thought of as a fine tuning or more precise determination of the speech detection achieved at the first stage. This point has been made at least in part by the above remarks pertaining to the rejection under Section 112. As explained

thereat, and as recited, e.g., in the "wherein" clause of the independent claims, the two stages (e.g., the speech detection means, on the one hand, and the likelihood determining means and the location determining means, on the other hand) are not entirely independent of one another. The second stage operates only as a part of the first stage. For example, once speech is detected in the first stage, the second stage operates on the N most recent frame energies, i.e., the second stage operates on a particular region of the input speech signal that is determined by the first stage. It is also noted that, as exemplified in the specification, the identification of the presence of speech at the first stage may be performed by a different method than the starting point/endpoint location determination using the likelihood determinations, which is performed at the second stage. The means or steps of the first stage do not (necessarily) operate to perform a likelihood-based determination of the starting point/endpoint of speech.

Chigier relates to automatic speech recognition involving assigning boundary probabilities to received frames and adjusting the boundary probabilities. Roughly speaking, *Chigier* describes a neural network phoneme recognizer. The neural network identifies boundaries between phonemes. Some such boundaries correspond to boundaries between speech and background noise. *Chigier* also uses energy measures and boundary probabilities. However, *Chigier*'s method identifies boundaries using only a single stage, whereas the claimed invention involves two stages of speech detection, in which, e.g., the second stage performs a finer or more precise detection than the first. As conceded by the Office Action, *Chigier* does not teach or suggest speech detecting means distinct from likelihood determining means (and location determining means). Applicant submits that nothing in *Chigier* would suggest at least the above-noted features of the independent claims.

Gupta et al. relates to a voice activity detector for speech signals in variable background noise. *Gupta et al.* uses certain features of a speech signal to discriminate between speech and noise, specifically, the features of level, slope and zero crossing. Depending on the measured numerical values or magnitudes of these features of a (portion of a) speech signal, a voice activity flag (VAD) is set to 1 or 0, indicating the presence or absence of voice activity, or speech, respectively. In this process, two threshold values for the level, a low level threshold and a high level threshold, are employed in making the determination as to whether there is voice activity or not. (The statement at page 5 of the Office Action, "When a VAD flag is set to one, then speech is compared to a first threshold, and when a VAD flag is set to zero, then speech is compared to a second threshold," is not understood to be an accurate description of the operation of *Gupta et al.*). Like *Chigier*, *Gupta et al.* teaches VAD using only a single stage, not two-stage speech detection, in which, e.g., the second stage performs a finer or more precise detection than the first. Applicant submits that nothing in *Gupta et al.* would suggest at least the above-noted features of the independent claims.

The Office Action would appear to allege that incorporating the teachings of *Gupta et al.* in the invention of *Chigier* would yield Applicant's claimed invention. However, Applicant respectfully disagrees. Combining (the teachings of) two single-stage speech detectors will not yield the functionality of a two-stage detector. A two-stage detector is more than the sum of the parts of two single-stage detectors. Nothing in either *Chigier* or *Gupta et al.* suggests that combining the invention of the one with the invention of the other would yield the additional functionality of a two-stage detector. Any such suggestion is understood to be the result of hindsight, based on additional knowledge gleaned from Applicant's disclosure. M.P.E.P.

2145.X.A. Further, neither *Chigier* nor *Gupta et al.* suggests that the method of speech detection of the one could be successfully used together with the method of speech detection of the other.

Moreover, even if, for the sake of argument, it were assumed that the two different speech detectors of *Chigier* and *Gupta et al.* could be combined, the resulting combination still does not suggest the relationship between the two stages of speech detection expressed by the "wherein" clause of Applicant's independent claims. That is, even if (the teachings of) *Chigier*'s speech detector and *Gupta et al.*'s speech detector could be combined, the resulting combination would not include the feature that one stage of speech detection (based on a likelihood determination method) is restricted to operate only when speech is detected through the operation of the other stage of speech detection.

Since neither *Chigier* nor *Gupta et al.*, whether taken singly or in combination (even assuming, for the sake of argument, that such combination were permissible), is understood to teach or suggest all of the elements of any of Applicant's independent claims, those claims are believed allowable over those documents.

A review of the other art of record, including *Cohrs et al.*, *Lennig et al.*, *Abut et al.*, and *Erell et al.*, has failed to reveal anything which, in Applicant's opinion, would remedy the deficiencies of the art discussed above, as references against the independent claims herein. These claims are therefore believed patentable over the art of record.

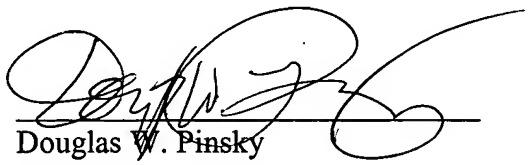
The other claims in this application are each dependent from one or another of the independent claims discussed above and are therefore believed patentable for the same reasons. Since each dependent claim is also deemed to define an additional aspect of the

invention, however, the individual reconsideration of the patentability of each on its own merits is respectfully requested.

In view of the foregoing amendments and remarks, Applicant respectfully requests favorable reconsideration and early passage to issue of the present application.

Applicant's undersigned attorney may be reached in our Washington office by telephone at (202) 530-1010. All correspondence should continue to be directed to our below listed address.

Respectfully submitted,



Douglas W. Pinsky
Attorney for Applicant
Registration No. 46,994

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3800
Facsimile: (212) 218-2200
DWP/klm/llp

DC_MAIN 246731v1